Available online **www.ijpras.com**

International Journal of Pharmaceutical Research & Allied Sciences, 2025, 14(1):67-79
https://doi.org/10.51847/gN3SGMVeqk

Original Article

ISSN : 2277-3657
CODEN(USA) : IJPRPM

What Can Molecular Dynamics Legitimately Claim in Pharmaceutical Research? A Critical Review of Sampling, Validation, and Mechanistic Interpretation

Kenji Tanaka[1]*, Yui Kobayashi[1], Takeshi Nakamura[2]

[1]Department of MD Simulation and Critical Appraisal, Faculty of Pharmaceutical Sciences, University of Tokyo, Tokyo, Japan.
[2]Department of Mechanistic Interpretation and Validation, Faculty of Pharmacy, Kyoto University, Kyoto, Japan.

***Email:** tanaka.kenji@eng.u-tokyo.ac.jp*

ABSTRACT

Molecular dynamics simulation has become central to pharmaceutical research because it can represent time-dependent molecular behaviour at spatial and temporal resolutions that are difficult to access experimentally. Its interpretive reach, however, is often extended beyond what a simulated trajectory can directly support. Structural persistence may be described as thermodynamic stability, incomplete sampling may be presented as a conformational ensemble, binding-pose retention may be treated as evidence of affinity, and correlated motions may be interpreted as causal mechanisms. This critical methodological review examines the evidential conditions under which such claims are defensible. A transparent, question-led search and appraisal approach was used to compare literature addressing conformational sampling, energy landscapes, force-field dependence, free-energy estimation, binding kinetics, allostery, convergence, reproducibility, and experimental validation. The appraisal focused on the proximity between computational output and stated conclusion, the adequacy of sampling, sensitivity to modelling assumptions, treatment of uncertainty, replication, external validation, and relevance to pharmaceutical decisions. The reviewed evidence indicates that molecular dynamics can robustly support bounded descriptions of behaviour produced by a specified model and, under stronger conditions, conditional inferences about conformational populations, thermodynamics, kinetics, and functional mechanisms. Confidence decreases when trajectory-local observations are converted into equilibrium, causal, or translational claims without independent replication, model-sensitivity analysis, quantity-matched validation, or experimental triangulation. The central judgement is that the legitimacy of a molecular-dynamics claim depends less on trajectory length or visual plausibility than on whether the evidence addresses the inferential step being made. Molecular dynamics is therefore most informative when claims remain proportional to sampling coverage, model fidelity, uncertainty, and external validation. The resulting framework is intended as a critical interpretive guide rather than a validated scoring system, clinical recommendation, regulatory standard, or universally applicable readiness framework.

Key words: *Molecular dynamics, Conformational ensemble, Energy landscape, Sampling, Drug–target recognition, Simulation validation*

Received: 12 November 2024; **Revised:** 20 February 2025; **Accepted:** 21 February 2025

INTRODUCTION

Molecular dynamics (MD) provides atomistic representations of molecular behaviour by propagating coordinates through time under defined modelling assumptions. In pharmaceutical research, MD is used to investigate protein

flexibility, ligand recognition, conformational transitions, solvent effects, and binding processes. However, trajectories are not direct recordings of molecular reality; they are simulations generated from selected force fields, initial conditions, boundary conditions, and sampling strategies. MD can therefore reveal consequences of a model but cannot independently establish that the simulated behaviour fully represents the biological system [1].

The interpretive challenge increases when MD is used beyond description as a computational assay of pharmaceutical activity. A trajectory may suggest stable binding, conformational change, or mechanistic hypotheses, but these interpretations require additional evidence. Structural observations do not automatically establish binding affinity, biological relevance, or therapeutic consequence. As claims progress from describing simulated behaviour toward prediction, mechanism, or decision support, the required evidential support must also increase [2].

Thermodynamic, kinetic, and mechanistic claims require particular caution because they describe different physical quantities. Binding affinity reflects equilibrium energetics, whereas kinetics concerns transition rates, pathways, and barriers. Similarly, preserved ligand contacts do not by themselves demonstrate free energy, residence time, or functional impact. These questions require appropriate sampling strategies, estimators, uncertainty analyses, and validation approaches [3].

Rare events such as association, dissociation, and allosteric transitions often require enhanced sampling, state models, or accelerated methods. Although these approaches expand accessible timescales, they introduce assumptions regarding sampling, reaction coordinates, state definitions, and time reconstruction. This review therefore examines what MD can legitimately claim in pharmaceutical research, focusing on boundaries between structural description, ensemble inference, prediction, mechanism, validation, and decision use [4].

Critical-review questions and evidential boundaries

This review evaluates six central questions: which MD claims require reassessment; which assumptions support interpretation; what evidence strengthens or weakens these assumptions; where modelling and validation practices create misleading conclusions; which claims remain defensible; and what evidence hierarchy follows from these distinctions. The primary unit of evaluation is the claim rather than the simulation method itself, because the same trajectory may support a limited description while failing to justify stronger mechanistic or predictive conclusions.

Four boundaries guide the analysis. The first separates model output from physical-system inference, as simulated behaviour depends on force fields, solvent models, parameters, and protocols. The second separates observed trajectory behaviour from ensemble properties, since finite sampling does not necessarily define equilibrium populations. The third separates association from causation, as correlated motions may generate hypotheses but do not prove functional control. The fourth separates computational performance from pharmaceutical utility, because predictive accuracy alone does not establish improved decision-making.

Evidence strength is therefore considered according to the inferential step being evaluated. Trajectory analysis may support descriptive claims, replicas may test robustness, sensitivity analyses may reveal dependence on modelling choices, and experimental comparisons may evaluate external consistency. These principles represent a proposed critical framework rather than fixed validation thresholds or regulatory criteria.

Literature identification and critical-appraisal approach

The literature was identified through a transparent, question-led search strategy combining MD with concepts including conformational ensembles, energy landscapes, enhanced sampling, force fields, free-energy calculations, binding kinetics, allostery, convergence, reproducibility, uncertainty, and experimental validation. Citation-chain searching was used to identify influential methodological studies, comparative analyses, and validation-focused research. Selection prioritised studies that supported specific methodological judgements or evidence boundaries rather than papers that merely used MD.

Eligible studies were peer-reviewed articles directly relevant to the review questions. Primary studies were prioritised for evaluating force fields, sampling methods, free-energy calculations, kinetic modelling, reproducibility, and mechanistic interpretation, while reviews supported broader conceptual synthesis. Studies lacking direct relevance or requiring conclusions beyond their examined system, method, or validation context were excluded. The review does not claim exhaustive coverage.

Appraisal focused on the relationship between computational evidence and claimed conclusions. Articles were examined for sampling adequacy, sensitivity to modelling assumptions, uncertainty treatment, replication, experimental comparison, and external validity. Evidence was interpreted comparatively: successful applications were not considered universal validation, and methodological limitations were not interpreted as eliminating the

value of MD. The synthesis instead evaluates which claims remain defensible after assumptions, uncertainties, and alternative explanations are considered.

Claims based on structural stability and conformational sampling

Structural claims from MD often rely on RMSD, hydrogen bonds, ligand poses, residue fluctuations, or conformational clusters. These describe sampled behaviour but do not automatically establish equilibrium states or complete sampling. State-based approaches such as Markov state models can identify metastable states and transitions, but their conclusions depend on state definitions, sampling depth, lag-time assumptions, and validation [5].

Enhanced sampling and generative approaches can expand exploration of conformational space, but broader sampling does not guarantee correct equilibrium populations. Generated structures must be distinguished from validated ensembles because accuracy depends on training data, reweighting, and representation of relevant states [6].

Sampling quality also depends on the molecular model itself. Force fields may favour particular conformations and perform differently across folded proteins, disordered regions, ligands, and solvent environments. Improved force fields represent progress but also highlight that predicted populations may reflect model assumptions as well as physical behaviour [7].

Therefore, the strongest structural claim is limited to behaviour observed under defined simulation conditions. Claims about dominant states, equilibrium populations, or pharmaceutical relevance require replication, convergence analysis, sensitivity testing, and external validation [8].

Claims concerning binding affinity, kinetics, and free energy

Binding affinity, kinetics, and free energy describe distinct physical properties and cannot be inferred interchangeably from a trajectory. Free-energy calculations can provide useful quantitative predictions when sampling, thermodynamic pathways, uncertainty estimation, and validation are appropriate. However, successful predictions in defined systems do not establish universal accuracy across targets or chemical spaces [9].

Large-scale drug-discovery assessments show that relative binding free-energy methods may support lead optimization, but performance depends on chemical series, protocol choices, binding modes, and model assumptions. Their value is therefore workflow-dependent rather than universally transferable [10].

Binding kinetics addresses transition pathways and rates rather than equilibrium affinity. Residence time cannot be inferred from pose stability, contact persistence, or docking scores alone. It may provide useful information in selected systems but is not a universal surrogate for efficacy or therapeutic duration [11].

Accelerated unbinding methods can estimate comparative kinetic behaviour but rely on assumptions about acceleration, pathways, starting states, and reconstruction of physical time. They support conditional kinetic hypotheses when externally validated but do not independently establish universal dissociation mechanisms or rates [12].

Table 1 organizes the evidence, constructs, relationships, uncertainties, and boundary conditions required for contested claims, hidden assumptions, and evidence standards in what can molecular dynamics legitimately claim in pharmaceutical research.

Table 1. Contested claims, hidden assumptions, and evidence standards in What Can Molecular Dynamics Legitimately Claim in Pharmaceutical Research?

Evidence domain or review construct	Review question	Eligible evidence type	Key extraction fields	Appraisal or relevance criterion	Synthesis role	Main limitation	Interpretation boundary
Trajectory-local structural stability	What does persistence within a finite trajectory directly establish?	Replicated conventional MD, trajectory diagnostics, structural-observable comparisons	Starting structure, simulation length, replicas, RMSD definition, contacts,	Whether the reported conclusion remains limited to sampled behaviour	Distinguishes direct observation from equilibrium inference	A plateau may reflect kinetic trapping, restricted coordinates, or model bias	"Remained stable in the sampled trajectory" does not establish experimental or

			restraints, equilibration				thermodynamic stability
Conformational ensemble	When can sampled conformations represent an equilibrium distribution?	State models, enhanced sampling, equilibrium-sampling methods, ensemble experiments	State definition, transition coverage, populations, lag time, reweighting, uncertainty	Evidence that relevant states and transitions are connected and population estimates are robust	Tests whether structural diversity supports an ensemble claim	Unvisited states and state-definition choices may alter the inferred landscape [5]	Generated or clustered conformations do not alone prove a complete ensemble
Force-field fidelity	Could the observed behaviour arise from the molecular model rather than the physical system?	Force-field development, comparative benchmarking, multi-observable validation	Parameter set, solvent model, protein class, ligand chemistry, calibration observables	Performance across relevant states and observables rather than one structural metric	Separates sampling uncertainty from Hamiltonian error	Calibration domains remain finite and may favour particular conformations [7]	Improved agreement does not establish universal physical fidelity
Binding affinity and selectivity	Does the calculation estimate a defined equilibrium quantity?	Absolute or relative alchemical free-energy studies with external measurements	Thermodynamic cycle, transformations, state preparation, convergence, uncertainty, experimental comparator	Quantity-matched validation within a defined chemical and target domain	Supports bounded quantitative prediction claims	Protonation, binding-mode, water, and reorganization errors may dominate [9]	A calculated free energy is not a measured affinity and cannot be generalized outside its tested domain
Pharmaceutical free-energy workflow	Does prediction performance translate into reliable compound prioritization?	Multi-project or prospective discovery assessments	Chemical-series coverage, failed transformations, uncertainty, prospective use, decision context	Stability of performance across projects and explicit handling of failures	Connects computational accuracy to decision relevance	Aggregate accuracy may conceal project-specific failure [10]	Workflow utility does not establish downstream therapeutic success
Binding kinetics and residence time	Is a physical rate or only a simulation-derived kinetic proxy being estimated?	Rare-event simulations, accelerated unbinding, state models, experimental kinetic comparison	Pathway assumptions, acceleration, transition definitions, rate reconstruction, (k_{on}), (k_{off})	Agreement with the corresponding experimental kinetic quantity	Separates kinetic prediction from pose persistence and affinity	Rare pathways and biased dynamics may distort reconstructed time [12]	A retained pose does not establish residence time
Allosteric or functional mechanism	Does dynamic association support causal functional interpretation?	MD integrated with structural, mutational, spectroscopic, or functional evidence	State dependence, pathway definition, perturbation, competing mechanisms, functional endpoint	Whether predicted coupling survives discriminating perturbation	Evaluates progression from correlation to mechanism	Multiple networks may explain the same correlated motions	Correlation or temporal ordering alone does not establish allosteric causation
Validation and decision use	What evidence is required for an intended pharmaceutical claim?	Independent experimental tests, prospective benchmarks,	Intended use, independent data, applicability domain,	Match between validation evidence and the exact	Defines consequence-scaled evidence requirements	Reproducibility and agreement may coexist	Technical performance does not automatically establish

	replication, sensitivity analysis	uncertainty, failure conditions	claimed quantity	with shared model bias	clinical, regulatory, or pharmaceutical utility

Mechanistic claims involving allostery and functional consequences

Mechanistic interpretation seeks more than a description of motion. It proposes that particular states, interactions, or transitions explain how a molecular perturbation produces a functional consequence. In G-protein-coupled receptors, molecular dynamics has been especially influential because available structures capture only selected receptor states, whereas signalling involves transitions among multiple conformations. Simulations can identify intermediate configurations, microswitch rearrangements, hydration changes, ion movements, and correlated structural transitions consistent with activation or deactivation. These observations can organize structural evidence into plausible dynamic narratives, but they remain hypotheses about mechanism unless the predicted state sequence or coupling relationship is tested independently [13].

Allostery presents a related challenge. An allosteric mechanism involves communication between spatially separated regions such that perturbation at one site changes energetic populations or functional behaviour elsewhere. Structural studies, pharmacological data, mutagenesis, and receptor-state comparisons provide evidence that such coupling occurs in many receptor systems. Molecular dynamics can propose how this coupling is transmitted through residues, helices, water networks, lipids, or ions. The inference becomes stronger when predicted pathways align with experimentally identified microswitches or when perturbing a proposed node alters the expected functional response. Structural connectivity or correlated movement alone, however, is insufficient because proteins contain many correlated fluctuations that may be consequences rather than drivers of functional change [14].

Network analysis, mutual information, covariance measures, community decomposition, and pathway algorithms can summarize relationships within high-dimensional trajectories. These tools are valuable for generating experimentally testable hypotheses, but their outputs depend on the selected structural features, correlation metric, threshold, simulated state, trajectory length, and treatment of indirect associations. Different analytical constructions may identify different pathways from the same underlying data. Moreover, statistical coupling does not establish the direction of information transfer or show that disrupting a high-centrality residue will alter function. MD-derived allosteric networks should therefore be interpreted as conditional representations of dynamic association, not direct maps of causal communication [15].

Large-scale analysis across receptor families can increase confidence when similar dynamic features recur across independently simulated systems. Recurring lateral gateways, internal cavities, or candidate allosteric regions may reveal structural principles that are difficult to recognize from isolated trajectories. Such recurrence supports prioritization for receptor-specific experiments, ligand design, or mutational testing. It does not establish that every identified site is pharmacologically accessible, functionally controlling, or therapeutically useful. Even broad comparative analyses remain conditioned by structural selection, simulation setup, receptor-state representation, and shared modelling assumptions. The appropriate conclusion is that these studies identify candidate allosteric sites and dynamic hypotheses requiring independent functional validation [16].

Convergence, replication, and sensitivity to modelling assumptions

A single long trajectory may appear persuasive because it contains substantial temporal detail, yet stochastic molecular dynamics can produce apparently meaningful differences that do not recur when initial velocities or starting microstates are changed. Independent replicas provide information that continuous extension of one trajectory cannot: whether a conclusion is robust to stochastic initialization and whether between-run variability exceeds the difference attributed to a ligand, mutation, or condition. Replica-based analyses have demonstrated that comparisons based on individual simulations can generate false-positive conclusions when variation among independent runs is ignored [17]. **Figure 1** presents the original conceptual synthesis for hierarchy of molecular-dynamics claims from descriptive to mechanistic, showing how the article's principal components, evidence relationships, uncertainties, and decision boundaries are connected.

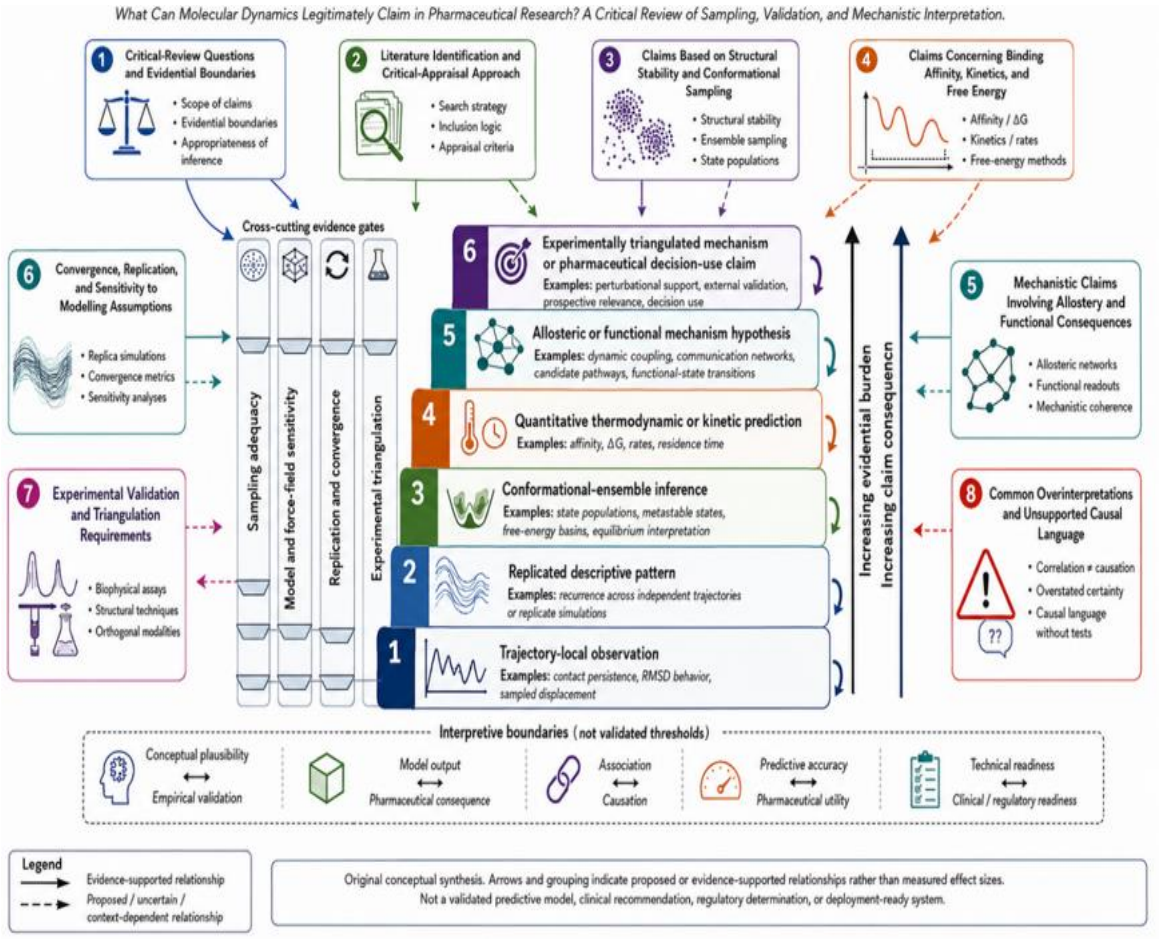

Figure 1. Hierarchy of Molecular-Dynamics Claims from Descriptive to Mechanistic

Reproducibility introduces a second dimension. Independent research groups may implement nominally equivalent models while making different choices about topology generation, equilibration, constraint algorithms, electrostatics, cutoffs, integration, thermostats, barostats, precision, or analysis. Round-robin simulations have shown that these procedural differences can produce meaningful variation even when participants work from a common scientific task [18]. Such variation does not necessarily indicate misconduct or software failure; it reveals that method descriptions often leave scientifically consequential choices implicit. Reproducibility therefore requires sufficient procedural detail to reconstruct not merely the conceptual method but the actual simulated system.

Cross-software comparisons provide a narrower test of implementation dependence. Agreement across simulation packages can increase confidence that a result is not an artefact of one code base, numerical convention, or default setting. Free-energy calculations reproduced across different packages have demonstrated that close agreement is achievable when molecular models, protocols, and estimators are carefully harmonized [19]. This is an important form of procedural verification. It does not, however, establish that the shared force field, chemical model, sampled states, or experimental interpretation is correct. Reproducibility can eliminate implementation-specific disagreement while preserving a common physical bias.

Sensitivity to molecular-model assumptions is distinct from stochastic and implementation uncertainty. Alternative force fields can assign different relative energies to backbone conformations, side-chain rotamers, folded states, disordered states, and ligand interactions. Comparative analysis of alanine conformational dynamics has shown that force-field choice can materially alter sampled distributions and their agreement with spectroscopic evidence [20]. Such sensitivity is especially important when a mechanistic claim depends on a small population shift, a rare transition, or a particular orientation. Replicas performed with one force field quantify stochastic variability within that model; they do not quantify uncertainty arising from plausible alternative models.

Convergence should therefore be evaluated at the level of the claimed conclusion. Stable temperature, energy, or RMSD does not demonstrate convergence of state populations, free-energy differences, transition pathways, kinetic rates, or allosteric networks. Observable convergence requires showing that the estimate and its uncertainty

no longer change materially under additional sampling within the intended model. State-space convergence asks whether relevant regions have been discovered and revisited. Model robustness asks whether the interpretation survives plausible changes to force field, solvent, protonation, starting state, and analysis procedure. Experimental validity asks whether the resulting prediction agrees with independent physical evidence. No single diagnostic answers all four questions.

Experimental validation and triangulation requirements

Experimental validation is frequently invoked as though it were a single binary event. In practice, simulations and experiments often report different representations of the system. A trajectory contains model-generated microscopic configurations, whereas experiments may report ensemble-averaged, time-averaged, spatially averaged, indirectly inferred, or condition-specific observables. Agreement therefore requires a forward model linking simulated configurations to the measured quantity and an account of uncertainty in both sources. Experiments should not be treated as error-free snapshots, but simulation–experiment complementarity can constrain interpretations that neither approach could resolve alone [21].

Kinetic validation is strongest when the simulated and experimental quantities are directly matched. Extensive molecular dynamics combined with state modelling has reconstructed protein–protein association pathways and yielded macroscopic kinetic estimates for a defined system [22]. This type of analysis is more compelling than identifying a visually plausible pathway because it connects microscopic transitions with an experimentally testable observable. Nevertheless, single-system success does not establish broad transferability. Similar performance in other systems may require different state decompositions, sampling depths, encounter-complex representations, or treatments of long-range interactions.

Experimental information can also be incorporated during model construction. Augmented Markov models provide a formal route for combining simulation trajectories with experimental observables to refine state populations or kinetic models [23]. This may improve consistency with available data and reduce uncertainty in poorly sampled quantities. A crucial distinction remains: evidence used to constrain a model cannot simultaneously function as fully independent validation of that model. Stronger evaluation requires withheld observables, independent experiments, perturbations, or prospective predictions that were not used to determine the final ensemble or transition network.

Structural-ensemble validation is similarly non-unique. Sparse or averaged measurements may be compatible with multiple ensembles, particularly when the number of possible conformations greatly exceeds the number of independent observables. Integrative determination therefore requires explicit forward models, uncertainty estimates, regularization assumptions, and checks for overfitting [24]. An experimentally compatible ensemble is not necessarily the uniquely correct ensemble, and agreement with one structural observable does not establish correct kinetics or function. **Table 2** organizes the evidence, constructs, relationships, uncertainties, and boundary conditions required for evidence supporting and challenging major claims, including methodological weaknesses.

Table 2. Evidence supporting and challenging major claims, including methodological weaknesses.

Method, platform, or evidence category	Representative application	Reported strength	Validation context	Generalization concern	Contradictory evidence	Evidence gap	Review interpretation
Conventional atomistic MD	Structural persistence, local rearrangements, interaction networks	High-resolution description of model behaviour over sampled time	Comparison with compatible structural or dynamical observables	Rare states and slow transitions may remain unobserved	Replica variability can reverse apparent differences [17]	Independent replicas and alternative starting states	Strong for bounded trajectory descriptions; insufficient alone for equilibrium or causal claims
Markov state modelling	Metastable-state identification and transition-network construction	Connects distributed trajectories into state and kinetic representations	Internal validation, implied timescales, external observables	State definitions and Markov assumptions can shape conclusions	Alternative discretizations may yield different networks [5]	Independent state definitions and prospective kinetic tests	Supports conditional ensemble or kinetic inference when state construction

							and validation are explicit
Enhanced or learned sampling	Exploration of rugged energy landscapes and rare conformations	May reach states inaccessible to conventional simulation	Reweighting, recovery of known distributions, independent comparison	Collective-variable or training-domain dependence	Greater diversity does not prove correct populations [6]	Robust tests for undiscovered states and extrapolation	Extends exploration but does not remove the need for convergence and model validation
Absolute or relative free-energy calculation	Ligand selectivity and lead optimization	Can provide quantitative ranking in defined target and chemical domains	Quantity-matched affinity data and project-level benchmarking	Performance may deteriorate with binding-mode changes, protonation uncertainty, or large reorganization	Project heterogeneity limits universal claims [10]	Prospective evaluation of failure detection and applicability domains	Potentially decision-relevant within validated workflows, not universally predictive
End-point free-energy estimation	Rapid post-processing of protein–ligand trajectories	Computational efficiency and qualitative ranking in some series	Retrospective comparison with experimental affinities	Strong dependence on entropy treatment, dielectric assumptions, and sampled conformations	More rigorous methods may show different ranking performance [25]	Protocol-specific calibration and uncertainty characterization	Appropriate as a conditional estimate, not a direct measurement of binding free energy
Accelerated unbinding and rare-event kinetics	Comparative residence-time estimation	Makes otherwise inaccessible dissociation events observable	Correlation with experimental kinetic ordering	Biased trajectories may not reproduce physical timing or pathway probabilities	Affinity or pose retention can disagree with kinetics [11]	Cross-target validation and explicit acceleration-to-rate uncertainty	Supports bounded kinetic hypotheses when externally calibrated
Dynamic-network and allostery analysis	Candidate communication pathways and allosteric-site identification	Converts complex motions into interpretable coupling hypotheses	Structural, mutational, spectroscopic, and functional triangulation	Network topology depends on features, thresholds, and simulated state	Correlation cannot establish directional causation [15]	Perturbations that distinguish proposed pathways from rivals	Valuable for hypothesis generation; mechanism requires discriminating evidence
Integrative experiment–simulation modelling	Ensemble refinement and kinetic-state estimation	Combines complementary spatial and temporal information	Forward-model agreement and independent observables	Data used for construction are not independent validation	Multiple ensembles may fit the same sparse data [24]	Withheld tests, perturbations, and quantified model non-uniqueness	Strongest when experimental roles in calibration and validation are clearly separated
Cross-software or cross-group reproducibility	Verification of free-energy and thermophysical workflows	Detects implementation and procedural dependence	Harmonized inputs and independently executed protocols	Shared agreement can preserve common force-field bias	Force-field sensitivity can remain despite software agreement [20]	Reproducibility studies coupled to experimental validation	Necessary for auditability, insufficient for physical validity

Common overinterpretations and unsupported causal language

A frequent error is treating MM/PBSA or MM/GBSA outputs as direct measurements of binding affinity. These methods provide protocol-dependent estimates influenced by sampling, solvent models, entropy treatment,

flexibility, and system-specific assumptions. They may support comparative ranking under defined conditions but are not equivalent to experimental free energies [26].

Comparisons between end-point and alchemical methods show that different approximations support different levels of inference. Performance varies across chemical systems, and retrospective agreement does not guarantee prospective reliability [25].

Qualitative stability approaches provide another boundary. Thermal-titration MD can compare relative protein–ligand stability under defined conditions but does not directly establish affinity, potency, selectivity, residence time, or clinical relevance [27].

Causal language should therefore be used cautiously. A trajectory can identify correlations, candidate pathways, or mechanistic hypotheses, but claims that a residue controls or a pathway drives function require independent perturbational evidence. Validation must always be linked to a defined purpose and external criterion [28].

A Hierarchy of defensible molecular-dynamics claims

The most defensible MD claims are trajectory-level observations: specific contacts, conformations, or transitions observed under defined conditions. Stronger claims about replicated patterns or conformational ensembles require independent simulations, broader sampling, convergence assessment, and validation against relevant observables.

Thermodynamic and kinetic predictions require additional evidence. Free-energy calculations should be evaluated against matched experimental quantities, uncertainty estimates, applicability domains, and potential failure modes rather than overall accuracy alone [29]. Energy landscapes are reconstructed models whose interpretation depends on coordinates, sampling, and analysis assumptions [30].

Machine-learning approaches introduce further uncertainty through training data, representations, and extrapolation limits. Improved sampling or computational efficiency does not automatically establish mechanistic validity or experimental accuracy [31].

The highest level of inference is pharmaceutical decision relevance. At this level, methods require validated workflows, uncertainty reporting, applicability limits, and evidence that outputs improve defined decisions. Free-energy calculations may support compound prioritization within controlled discovery settings but remain dependent on domain-specific validation [32].

Evidence gaps and reporting priorities

Reliable MD interpretation requires transparent reporting of simulation inputs, parameters, force fields, water models, equilibration procedures, production settings, and analysis methods. Reproducibility enables evaluation but does not alone establish physical validity [33].

Enhanced-sampling studies require additional reporting of biasing methods, collective variables, convergence assessment, reweighting procedures, and limitations. Method names alone are insufficient to evaluate whether sampled states and reconstructed quantities are reliable [34].

Figure 2 presents the original conceptual synthesis for evidence requirements for increasingly consequential pharmaceutical claims, showing how the article's principal components, evidence relationships, uncertainties, and decision boundaries are connected.

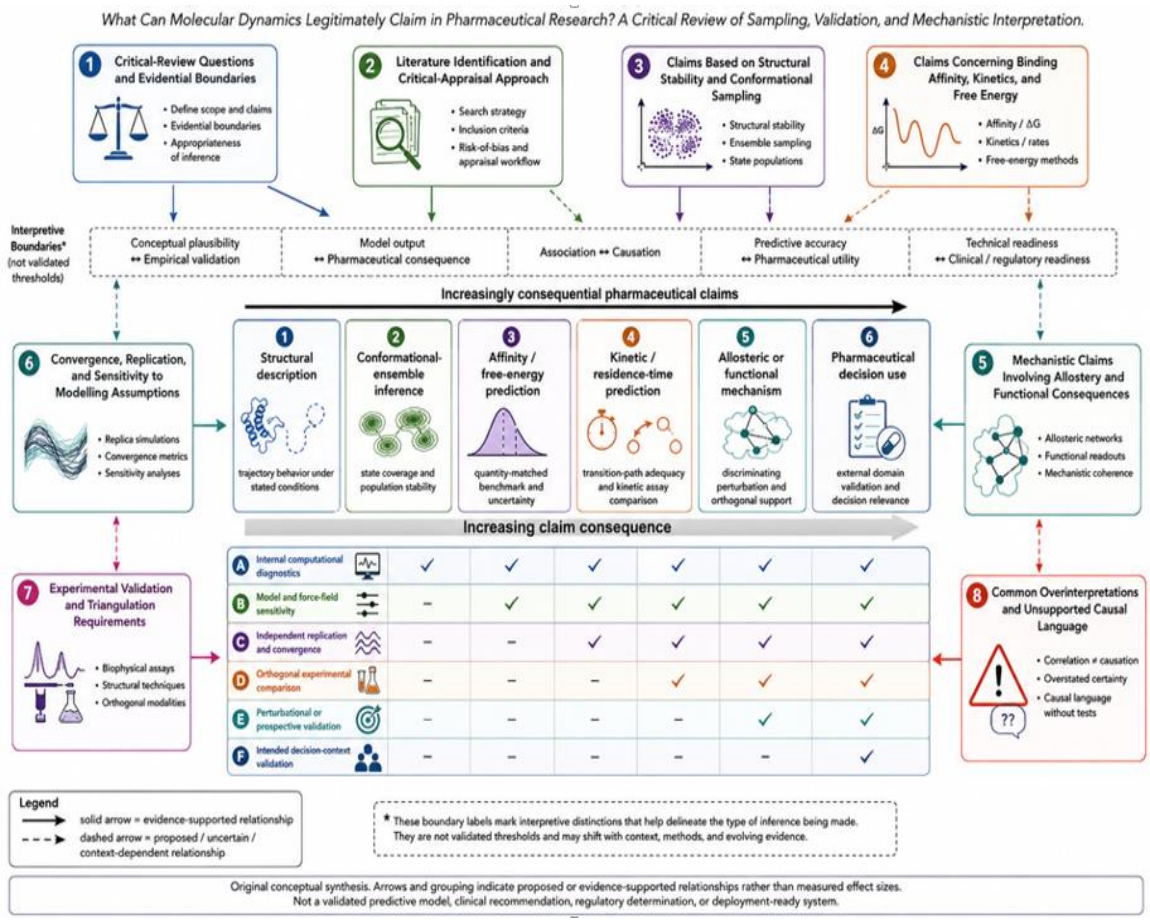

Figure 2. Evidence Requirements for Increasingly Consequential Pharmaceutical Claims.

Force-field validation also requires broader criteria. Structural agreement may be assessed through backbone distributions, side-chain preferences, folded-state stability, disordered-state dimensions, conformational populations, or agreement with spectroscopic and thermodynamic observables. These criteria differ in their ability to discriminate among parameter sets and may favour different models. Comparative validation against multiple structural criteria has shown why no single observable should be treated as sufficient evidence of overall force-field quality [35]. Reporting priorities should therefore include the intended domain of the force field, known calibration data, observables used for evaluation, and behaviours not tested.

The solvent model should receive comparable attention. Water influences hydrophobic association, hydrogen bonding, electrostatic screening, protein compactness, ligand desolvation, folding, and transition barriers. Comparative studies indicate that the water model can alter folding and unfolding dynamics and can interact with the nominal protein force field rather than acting as a neutral implementation detail [36]. Future studies should separate stochastic uncertainty, parameter uncertainty, structural-state uncertainty, and experimental uncertainty; report failed or unstable calculations; examine sensitivity to plausible molecular models; and predefine the language supported by each analysis. These practices would not guarantee correctness, but they would make claims more auditable and reduce the risk that apparently precise results conceal unexamined assumptions.

Limitations of the review

This review used a transparent, question-led critical approach rather than an exhaustive systematic design. The selected literature was chosen to evaluate major MD claim types, assumptions, validation challenges, and counterarguments. It should therefore be interpreted as a reasoned appraisal of claim defensibility rather than a quantitative assessment of errors across the field.

The evidence base varies across molecular systems, ligands, simulation protocols, force fields, software, and experimental comparators. A method validated in one context may not generalize to others with different chemical or conformational features. Validation must therefore remain linked to the intended purpose and tested domain [28].

Reproducibility, force-field evaluation, and experimental agreement address different uncertainty sources. Transparent protocols improve auditability but cannot correct inaccurate physical models, while structural validation does not necessarily establish correct kinetics or biological function. Current guidance supports improved reporting and evaluation but does not define universal thresholds for mechanistic or pharmaceutical validity [33, 35].

CONCLUSION

Molecular dynamics can legitimately claim that a defined molecular model produces specific behaviour under stated simulation conditions. Stronger claims require additional evidence addressing sampling, validation, uncertainty, and external relevance. Structural persistence is not equivalent to stability; observed conformations are not automatically equilibrium ensembles; retained poses are not affinity or residence time; correlated motions are not causal mechanisms; and computational performance is not pharmaceutical utility.

Replication, convergence analysis, sensitivity testing, uncertainty estimation, experimental comparison, perturbation studies, and decision-context validation answer different questions and should not be treated as interchangeable. The central contribution of this critical review is a claim hierarchy in which the strength of interpretation remains proportional to the available evidence. MD remains highly valuable as a source of mechanistic hypotheses, conditional predictions, and experimentally testable explanations, provided that its conclusions remain bounded by what the simulation and validation evidence can actually support.

ACKNOWLEDGMENTS: None

CONFLICT OF INTEREST: None

FINANCIAL SUPPORT: None

ETHICS STATEMENT: None

REFERENCES

1. Hollingsworth SA, Dror RO. Molecular dynamics simulation for all. Neuron. 2018;99(6):1129-43. doi:10.1016/j.neuron.2018.08.011
2. Huggins DJ, Biggin PC, Dämgen MA, Essex JW, Harris SA, Henchman RH, et al. Biomolecular simulations: From dynamics and mechanisms to computational assays of biological activity. Wiley Interdiscip Rev Comput Mol Sci. 2019;9(3). doi:10.1002/wcms.1393
3. Decherchi S, Cavalli A. Thermodynamics and kinetics of drug-target binding by molecular simulation. Chem Rev. 2020;120(23):12788-833. doi:10.1021/acs.chemrev.0c00534
4. Nunes-Alves A, Kokh DB, Wade RC. Recent progress in molecular simulation methods for drug binding kinetics. Curr Opin Struct Biol. 2020;64:126-33. doi:10.1016/j.sbi.2020.06.022
5. Husic BE, Pande VS. Markov state models: From an art to a science. J Am Chem Soc. 2018;140(7):2386-96. doi:10.1021/jacs.7b12191
6. Noé F, Olsson S, Köhler J, Wu H. Boltzmann generators: Sampling equilibrium states of many-body systems with deep learning. Science. 2019;365(6457). doi:10.1126/science.aaw1147
7. Robustelli P, Piana S, Shaw DE. Developing a molecular dynamics force field for both folded and disordered protein states. Proc Natl Acad Sci U S A. 2018;115(21). doi:10.1073/pnas.1800690115
8. Huang J, Rauscher S, Nawrocki G, Ran T, Feig M, de Groot BL, et al. CHARMM36m: An improved force field for folded and intrinsically disordered proteins. Nat Methods. 2017;14(1):71-3. doi:10.1038/nmeth.4067
9. Aldeghi M, Heifetz A, Bodkin MJ, Knapp S, Biggin PC. Predictions of ligand selectivity from absolute binding free energy calculations. J Am Chem Soc. 2017;139(2):946-57. doi:10.1021/jacs.6b11467
10. Schindler CEM, Baumann H, Blum A, Böse D, Buchstaller HP, Burgdorf L, et al. Large-scale assessment of binding free energy calculations in active drug discovery projects. J Chem Inf Model. 2020;60(11):5457-74. doi:10.1021/acs.jcim.0c00900
11. Bernetti M, Masetti M, Rocchia W, Cavalli A. Kinetics of drug binding and residence time. Annu Rev Phys Chem. 2019;70:143-71. doi:10.1146/annurev-physchem-042018-052340

12. Kokh DB, Amaral M, Bomke J, Grädler U, Musil D, Buchstaller HP, et al. Estimation of drug-target residence times by τ-random acceleration molecular dynamics simulations. J Chem Theory Comput. 2018;14(7):3859-69. doi:10.1021/acs.jctc.8b00230
13. Latorraca NR, Venkatakrishnan AJ, Dror RO. GPCR dynamics: Structures in motion. Chem Rev. 2017;117(1):139-55. doi:10.1021/acs.chemrev.6b00177
14. Thal DM, Glukhova A, Sexton PM, Christopoulos A. Structural insights into G-protein-coupled receptor allostery. Nature. 2018;559(7712):45-53. doi:10.1038/s41586-018-0259-z
15. Ribeiro JML, Filizola M. Allostery in G protein-coupled receptors investigated by molecular dynamics simulations. Curr Opin Struct Biol. 2019;55:121-8. doi:10.1016/j.sbi.2019.03.016
16. Aranda-García D, Stepniewski TM, Torrens-Fontanals M, García-Recio A, Lopez-Balastegui M, Medel-Lacruz B, et al. Large scale investigation of GPCR molecular dynamics data uncovers allosteric sites and lateral gateways. Nat Commun. 2025;16(1):2020. doi:10.1038/s41467-025-57034-y
17. Knapp B, Ospina-Forero L, Deane CM. Avoiding false positive conclusions in molecular simulation: The importance of replicas. J Chem Theory Comput. 2018;14(12):6127-38. doi:10.1021/acs.jctc.8b00391
18. Schappals M, Mecklenfeld A, Kröger L, Botan V, Köster A, Stephan S, et al. Round robin study: Molecular simulation of thermodynamic properties from models with internal degrees of freedom. J Chem Theory Comput. 2017;13(9):4270-80. doi:10.1021/acs.jctc.7b00489
19. Loeffler HH, Bosisio S, Duarte Ramos Matos G, Suh D, Roux B, Mobley DL, et al. Reproducibility of free energy calculations across different molecular simulation software packages. J Chem Theory Comput. 2018;14(11):5567-82. doi:10.1021/acs.jctc.8b00544
20. Zhang S, Schweitzer-Stenner R, Urbanc B. Do molecular dynamics force fields capture conformational dynamics of alanine in water? J Chem Theory Comput. 2020;16(1):510-27. doi:10.1021/acs.jctc.9b00588
21. Bottaro S, Lindorff-Larsen K. Biophysical experiments and biomolecular simulations: A perfect match? Science. 2018;361(6400):355-60. doi:10.1126/science.aat4010
22. Plattner N, Doerr S, De Fabritiis G, Noé F. Complete protein-protein association kinetics in atomic detail revealed by molecular dynamics simulations and Markov modelling. Nat Chem. 2017;9(10):1005-11. doi:10.1038/nchem.2785
23. Olsson S, Wu H, Paul F, Clementi C, Noé F. Combining experimental and simulation data of molecular processes via augmented Markov models. Proc Natl Acad Sci U S A. 2017;114(31):8265-70. doi:10.1073/pnas.1704803114
24. Bonomi M, Heller GT, Camilloni C, Vendruscolo M. Principles of protein structural ensemble determination. Curr Opin Struct Biol. 2017;42:106-16. doi:10.1016/j.sbi.2016.12.004
25. Aldeghi M, Bodkin MJ, Knapp S, Biggin PC. Statistical analysis on the performance of molecular mechanics Poisson-Boltzmann surface area versus absolute binding free energy calculations: Bromodomains as a case study. J Chem Inf Model. 2017;57(9):2203-21. doi:10.1021/acs.jcim.7b00347
26. Wang E, Sun H, Wang J, Wang Z, Liu H, Zhang JZH, et al. End-point binding free energy calculation with MM/PBSA and MM/GBSA: Strategies and applications in drug design. Chem Rev. 2019;119(16):9478-508. doi:10.1021/acs.chemrev.9b00055
27. Pavan M, Menin S, Bassani D, Sturlese M, Moro S. Qualitative estimation of protein-ligand complex stability through thermal titration molecular dynamics simulations. J Chem Inf Model. 2022;62(22):5715-28. doi:10.1021/acs.jcim.2c00995
28. van Gunsteren WF, Daura X, Hansen N, Mark AE, Oostenbrink C, Riniker S, et al. Validation of molecular simulation: An overview of issues. Angew Chem Int Ed Engl. 2018;57(4):884-902. doi:10.1002/anie.201702945
29. Mobley DL, Gilson MK. Predicting binding free energies: Frontiers and benchmarks. Annu Rev Biophys. 2017;46:531-58. doi:10.1146/annurev-biophys-070816-033654
30. Bussi G, Laio A. Using metadynamics to explore complex free-energy landscapes. Nat Rev Phys. 2020;2(4):200-12. doi:10.1038/s42254-020-0153-0
31. Noé F, Tkatchenko A, Müller KR, Clementi C. Machine learning for molecular simulation. Annu Rev Phys Chem. 2020;71:361-90. doi:10.1146/annurev-physchem-042018-052331
32. Cournia Z, Allen B, Sherman W. Relative binding free energy calculations in drug discovery: Recent advances and practical considerations. J Chem Inf Model. 2017;57(12):2911-37. doi:10.1021/acs.jcim.7b00564

33. Elofsson A, Hess B, Lindahl E, Onufriev A, van der Spoel D, Wallqvist A. Ten simple rules on how to create open access and reproducible molecular simulations of biological systems. PLoS Comput Biol. 2019;15(1). doi:10.1371/journal.pcbi.1006649
34. Bonomi M, Bussi G, Camilloni C, Tribello GA, Banáš P, Barducci A, et al. Promoting transparency and reproducibility in enhanced molecular simulations. Nat Methods. 2019;16(8):670-3. doi:10.1038/s41592-019-0506-8
35. Stroet M, Setz M, Lee T, Malde AK, van den Bergen G, Sykacek P, et al. On the validation of protein force fields based on structural criteria. J Phys Chem B. 2024;128(19):4602-20. doi:10.1021/acs.jpcb.3c08469
36. Fischer ALM, Tichy A, Kokot J, Hoerschinger VJ, Wild RF, Riccabona JR, et al. The role of force fields and water models in protein folding and unfolding dynamics. J Chem Theory Comput. 2024;20(5):2321-33. doi:10.1021/acs.jctc.3c01106